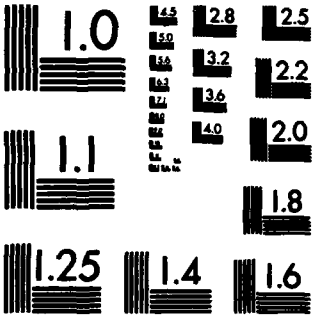
AD-A145 675 LIKELIHOOD ESTIMATION FOR GENERALIZED MIXED EXPONENTIAL DISTRIBUTIONS..(U) VIRGINIA UNIV CHARLOTTESVILLE SCHOOL OF ENGINEERING AND APPLI.. C M HARRIS ET AL. JUL 84 1/1

UNCLASSIFIED UVA/525393/SE85/104 N00014-83-K-0624 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

A Report

LIKELIHOOD ESTIMATION FOR GENERALIZED MIXED EXPONENTIAL DISTRIBUTIONS

Submitted to:

Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Attention: Group Leader, Statistics and Probability
Mathematical and Physical Sciences

Submitted by:

Carl M. Harris
Professor

Edward A. Sykes
Graduate Research Assistant

Report No. UVA/525393/SE85/104
July 1984

SCHOOL OF ENGINEERING AND
APPLIED SCIENCE

DEPARTMENT OF SYSTEMS ENGINEERING

UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA 22901

LIKELIHOOD ESTIMATION FOR GENERALIZED MIXED EXPONENTIAL DISTRIBUTIONS

Submitted to:

Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Attention: Group Leader, Statistics and Probability
Mathematical and Physical Sciences

Submitted by:

Carl M. Harris
Professor

Edward A. Sykes
Graduate Research Assistant

Department of Systems Engineering
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA

Report No. UVA/525393/SE85/104
July 1984

Copy No. ____

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Likelihood Estimation of Generalized Mixed Exponential Distributions		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER UVA/525393/SE85/104
7. AUTHOR(s) Carl M. Harris Edward A. Sykes		8. CONTRACT OR GRANT NUMBER(s) N00014-83-K-0624
9. PERFORMING ORGANIZATION NAME AND ADDRESS Systems Engineering, Thornton Hall University of Virginia Charlottesville, VA 22901		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 N. Quincy Street Arlington, VA 22217		12. REPORT DATE July 1984 13. NUMBER OF PAGES 28
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) maximum likelihood statistical inference parameter estimating phase-type distributions mixed exponential distribution nonlinear programming		

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The class of probability functions expressed as linear (not necessarily convex) combinations of negative exponential densities is dense in the set of all distribution functions on the nonnegative reals. Because of this and resultant mathematical properties, such forms would appear to have excellent potential for wide application in stochastic modeling. This work documents the development and testing of a practical procedure for maximum-likelihood estimation for these generalized exponential mixtures. The algorithm offered for the problem is of the Jacobi type and guarantees that the result will provide a legitimate probability

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (continued)

function of the prescribed type. Extensive testing has been performed and results are very favorable: convergence is rapid and the use of computer resources rather limited.

I. INTRODUCTION

Finite mixtures of continuous distributions are being used in an increasing number of stochastic modeling efforts (see Harris, Kaylan and Maltz, 1982; Everitt and Hand, 1981). The most common distribution applied appears to be the mixed exponential, where the probability density function would be written as a convex linear combination of (say) K exponential subpopulations

$$b(t) = \sum_{i=1}^{K} \gamma_i \phi_i e^{-\phi_i t} \qquad (\sum_{i=1}^{K} \gamma_i = 1). \tag{1}$$

Though such models have very general application, there are some serious restrictions on their use. Most importantly, the functional form of this PDF requires that its data exhibit quite a strong monotone decreasing pattern since b(t) always lies somewhere between the K monotone subdensities. (The precise mathematical term is that it is completely monotone.) Though it is true that the mixed exponential (with possibly an open number of terms) has broad potential coverage, it is clearly poorly suited for fitting data with apparent modes and indeed any set of frequencies seeming to be nonmonotone decreasing.

But a variation on this mixed exponential theme can be adapted to arbitrary data sets. This is to permit the linear proportionality constants $\{\gamma_i\}$ in the mixed exponential model to be totally arbitrary in sign, while still, of course, requiring that the resultant PDF remain nonnegative over its domain. The establishment of a practical numerical procedure for doing

maximum-likelihood estimation for these more general mixed exponential models is the major subject of this work.

We refer to these nonconvex forms as GHs for generalized hyperexponentials. It has frequently been noted in the literature (e.g., in Neuts, 1981) that the GH class is dense in the set of all CDFs on the nonnegative reals. By this, we mean that there is always a GH which is as close as we may wish to an arbitrary CDF measured with respect to a metric suitably defined on the distribution space. It is also important to recognize that the general class of linear combinations of negative exponential functions is a very complete set of approximants in general function space. Perhaps the most critical characterization of this coverage is the fact that all functions in $L_2(0,\infty)$ can be approximated arbitrarily closely by a finite linear combination of functions of the form $e^{-\phi t}$, $\phi \varepsilon R^+$. (For example, see Naylor and Sell, 1971, for a discussion of this and related problems on the Hilbert space of square integrable functions.)

The GH class has important relationships with a number of well-known comprehensive families of exponentially related CDFs. The simplest of these are the generalized Erlangs (GE) expressed as convolutions of independent and not-necessarily-identical exponential random variables. When the means of such exponentials are allowed to come in conjugate pairs (so that their Laplace-Stieltjes transforms are inverse polynomials), Smith (1953) calls the family K_n, where n is the degree of the defining polynomial. Cox (1955) generalized K_n to the class of distributions whose transforms are rational functions (clearly including the inverse polynomials), which we call R_n (with n the degree of the denominator's polynomial). The K_n class includes all regular Erlangs, but not all mixed exponentials and mixed Erlangs, which are, however, members of R_n. We also mention the generalized phase-type

distributions (PH) popularized by Neuts and others (see Neuts, 1981), which have rational transforms as well, though not necessarily of the inverse polynomial form. Thus, we may symbolically represent the relationship of those respective families as $GE \subset K_n \subset R_n$ and $PH \subset R_n$.

To position the GH class together with the others, we first observe that any member of R_n with real and distinct zeroes for its transform denominator polynomial is a generalized hyperexponential. But not all GH are phase types, since there are linear combinations of exponentials which are densities but cannot be derived as the time to absorption of any Markov chain (see Dehon and Latouche, 1982).

The numerical procedure developed for estimating the parameters of the generalized hyperexponentials has been built up from previous work on exponential and Weibull mixtures. Throughout we assume that the data sampling is complete so that all random times are fully observed. In the event that there are incomplete data observations, the algorithm is easily altered.

Maximum-likelihood estimation is the method selected mainly because, under fairly general conditions, it enjoys the important limiting statistical properties of efficiency, normality, and unbiasedness. Furthermore, the MLEs are consistent, invariant, and are functions of sufficient statistics if they exist. When sufficiency and unbiasedness both hold, the MLEs are also of minimum variance.

A first key observation is that it is not possible to obtain explicit formulas for the maximum-likelihood estimators of parameters involved in mixed exponential densities by taking the partial derivatives and equating them to zero. Hence we resort to other optimization methods and numerical techniques. Furthermore, we need to take into account a set of constraints in addition to the objective function. The mixing proportions and scale parameters must

satisfy some simple linear relationships and there may exist other constraints related to the sub-population parameters. Note that the constraints are generally of a linear type; hence the problem can be described as a mathematical program with a nonlinear objective function and linear constraints.

II. PROBLEM STATEMENT

The target criterion function in our maximum-likelihood optimization problem is the usual joint density function for a random sample from the population governed by the b(t) of (1). As is common, it is much easier in this situation to work with the logarithm of the likelihood function. Thus, if we write the likelihood for the random sample $t_1,\ldots,t_N$ as

$$\mathcal{L}(\alpha) = \prod_{j=1}^{N} b(t_j;\alpha) = \prod_{j=1}^{N} \sum_{i=1}^{K} \gamma_i \phi_i e^{-\phi_i t_j} \qquad (2)$$

where α is the vector of parameters (which may include K), then its logarithm is expressed as

$$L(\alpha) = \sum_{j=1}^{N} \ln b(t_j;\alpha) \quad . \qquad (3)$$

The MLE problem for the (generalized) mixture may then be formulated as the nonlinear constrained optimization problem:

$$\max_{\alpha} L(\alpha)$$

subject to

$$\alpha \in S = \{\alpha \mid \Sigma\gamma_i = 1; \phi_i \geq 0\} \quad . \qquad (4)$$

Under the standard mixed-exponential regime, each $\gamma_i > 0$, and ϕ_i would be real and also greater than 0. The most efficient algorithm available

for the solution of this problem is due to the joint efforts of Kaylan (1978) and Kaylan and Harris (1981), and Mandelbaum (1982) and Mandelbaum and Harris (1982). It is a sequential numerical procedure whose principal sequence of points is generated using nonlinear Jacobi-like iterations, but with an embedded subsequence generated by Armijo steps. Details will be provided below.

Because the mixing parameters $\{\gamma_i\}$ may be negative in our problem, the Kaylan/Mandelbaum/Harris (K/M/H) algorithm has been carefully altered, though its basic approach is retained. Three major changes in the algorithm were necessary. First, additional code had to be added to make sure that the density function b(t) did not become negative. The second and third alterations were required to ensure that the algorithm generates ascent directions for the $\{\phi_i\}$ and $\{\gamma_i\}$, respectively, when the sign restriction on the $\{\gamma_i\}$ is relaxed.

III. DESCRIPTION OF THE ALGORITHM

The basis of the algorithm is the first-order, nonpost-mortem method for the convex mixing of Weibull distributions described in Mandelbaum and Harris (1982). Exponential distributions are, of course, Weibull with shape parameter one. This numerical scheme calculates the parameter values α^{v+1} at the (v+1)st iteration as

$$\alpha^{v+1} = \alpha^{v} + s^{v}d(v) \quad (v = 0,1,2,\ldots), \tag{5}$$

where

α^{v} is the vector of parameter values at the (v)th iteration, i.e., α^{v} $= \{\phi_1^{v}, \phi_2^{v}, \ldots, \phi_K^{v}, \gamma_1^{v}, \gamma_2^{v}, \ldots, \gamma_{K-1}^{v}\}$ (γ_K obtains from the sum-to-one constraint);

$d(v)$ is the ascent direction generated at the (v)th iteration (described below);

$s^v = 2^{-i(v)}$ is the step size at the (v)th iteration; and $i(v)$ is the smallest nonnegative integer such that $L(\alpha^{v+1}) - L(\alpha^v) \geq \varepsilon(v) \geq 0$ ($\varepsilon(v)$ will be described below), $\alpha^{v+1} \in S$ and $b(t)$ is nonnegative.

The algorithm assumes α^0 is user-supplied and terminates at the first iteration v for which $i(v) = i_{max}$ (= 20, currently). That is, if twenty bisections of the step-size do not yield an α^{v+1} which is feasible, which forms a density $b(t)$, and which evaluates to an increase in the log-likelihood function, the algorithm terminates at the presumed local maximum α^v.

The ascent directions $d(v)$ are calculated in one of two ways, depending on the value of the iteration index v and a constant W (presently 40). If v is not a multiple of 40, the regular iteration is employed as a variation of the nonlinear Jacobi step (see Ortega and Rheinboldt, 1970) formed by making use of $\nabla L(\alpha) = 0$ (see Mandelbaum and Harris, 1982, for details). Otherwise, if v is a multiple of 40, then a gradient-based Armijo step is used.

The necessity for the two types of iterations arises in the development of the K/M/H algorithm from the use of step-size bisection in the mapping from α^v to α^{v+1}. Without bisection, i.e., if $s^v = 1$ for all v, the original algorithm's mapping (5) is closed since all functions involved are continuous, and hence α^v belongs to a compact set. But with bisection, continuity might occasionally be violated and convergence prevented. To avoid this potential pathology, the Armijo step, which is convergent under bisection (see Armijo, 1966) was included to embed a convergent subsequence of points into the principal sequence.

The potential violation of continuity may be further aggravated by allowing unrestricted signs on each γ_i in the current algorithm. Though empirically, the $\{\gamma_i\}$ appear not to change from their initial signs very often, in theory, the potential change in sign could cause discontinuity in the functions comprising the principal mapping and hence, hinder convergence of the principal sequence. The gradient-based Armijo step would not be subject to this discontinuity and hence, serves to provide a convergent subsequence of points for this case as well.

The ascent direction calculation of the regular iteration is where it has been necessary to alter the Kaylan/Mandelbaum/Harris algorithm. The original algorithm, at the (v)th iteration, is defined by solving the equations

$$\frac{\partial L}{\partial \alpha_i}(\alpha_1^v, \alpha_2^v, \ldots, \alpha_{i-1}^v, \alpha_i, \alpha_{i+1}^v, \ldots, \alpha_{2K-1}^v) = 0 \quad (i = 1,2,\ldots,2K-1) \qquad (6)$$

for each α_i, $i = 1,2,\ldots,2K-1$, in sequence, and setting $d(v) = \alpha - \alpha^v$. To allow the equations of (6) to be resolved as

$$\alpha_i = h_i(\alpha^v) \quad (i = 1,2,\ldots,2K-1), \qquad (7)$$

we allow α_i^v to be used on the right-hand side if a numerical procedure would have been needed to isolate α_i. The explicit form of the equations of (7) and the details of their derivation can be found in Mandelbaum (1982) and Mandelbaum and Harris (1982). It should be noted here that although the regular iteration makes use of gradient information, it is not a gradient step, since the equations (6) are solved sequentially rather than simultaneously. The gradient (steepest ascent) direction would be that derived through solving $dL/d\alpha = 0$ where α and 0 are vectors.

The modified regular iteration calculates the vector α as previously, but forms the ascent direction d(v) differently. Recalling that

$$\alpha = (\alpha_1 = \phi_1,\ \alpha_2 = \phi_2, \ldots, \alpha_K = \phi_K,\ \alpha_{K+1} = \gamma_1,\ \alpha_{K+2} = \gamma_2, \ldots, \alpha_{2K-1} = \gamma_{K-1}) \tag{8}$$

where K is the number of terms in b(t) (γ_K is always calculated as one minus the sum of the K-1 $\{\gamma_i\}$), the elements of the ascent direction are calculated as follows. The $d_i(v)$, i = 1,2,...,K, (i.e., the elements of the ascent direction associated with the exponential subpopulation shape parameters $\{\phi_i\}$) are

$$d_i(v) = \begin{cases} \alpha_i - \alpha_i^v & \text{if } \gamma_i \geq 0 \\ \alpha_i^v - \alpha_i & \text{if } \gamma_i < 0 \end{cases} \qquad (i = 1,2,\ldots,K) \tag{9}$$

To calculate the (K+1)st, (K+2)nd, through the (2K-1)th elements of $d_i(v)$, (i.e., the elements of the ascent direction associated with the mixing coefficients $\{\gamma_i,\ i = 1,2,\ldots,K-1\}$), we first calculate a test parameter s as

$$s = \sum_{i=1}^{K} \frac{(\gamma_i - \gamma_i^v)^2}{\gamma_i^v} . \tag{10}$$

Then

$$\begin{aligned} &\text{If } s \geq 0, \quad d_i(v) = \alpha_i - \alpha_i^v \ (i = K+1,K+2,\ldots,2K-1) \\ &\text{Else } (s < 0), \quad d_i(v) = \alpha_i^v - \alpha_i \ (i = K+1,K+2,\ldots,2K-1) \end{aligned} \tag{11}$$

The modifications in the ascent direction calculation for the regular iteration obtain from a slight extension of the ascent direction proof for the original algorithm (see Mandelbaum, 1982, or Mandelbaum and Harris, 1982). The sufficient condition for proof of ascent direction is

$$d(v) \cdot \nabla L(\alpha^v) \geq 0. \tag{12}$$

From a simple extension of the original proof, we find that

$$(\phi_i - \phi_i^v) \cdot \frac{\partial L(\alpha^v)}{\partial \phi_i} = \frac{\left[\frac{\partial L(\alpha^v)}{\partial \phi_i}\right]^2}{\gamma_i^v \, (\phi_i^v)^2 \sum_{j=1}^{N} \frac{\phi_i^v \, e - \phi_i^v \, t_j}{b(t_j)}} \qquad (i = 1,\ldots,K) \tag{13}$$

where N is the number of observed lifetimes. Examining the RHS of (13), since the squared terms are nonnegative and the terms of the summation are by definition nonnegative, the sign of the RHS is determined by the sign of γ_i^v. Hence, (13) or its negative will always have a RHS ≥ 0, and thus rule (9) is satisfied.

As with rule (11), the ascent direction equations from which (11) is derived handle the $\{\gamma_i,\ i = 1,\ldots,K-1\}$ together. Since under rule (9), the first K elements of the inner product (12) are nonnegative, the sufficient condition for proof of an ascent direction reduces to

$$\sum_{i=K+1}^{2K-1} d_i(v) \, \frac{\partial L(\alpha^v)}{\partial \alpha_i} \geq 0. \tag{14}$$

Once again, a simple extension of the original proof yields

$$s' = \sum_{i=1}^{K-1} [\gamma_i - \gamma_i^v] \, \frac{\partial L(\alpha^v)}{\partial \gamma_i} = N \sum_{i=1}^{K} \frac{(\gamma_i - \gamma_i^v)^2}{\gamma_i^v} \tag{15}$$

where N is the number of observations. Noting that in (10), $s = s'/N$, and since s' or -s' will always be nonnegative, rule (11) obtains directly.

The ascent direction calculation of the Armijo iteration of the new algorithm is identical to that of the original algorithm; that is,

$$d(v) = \nabla L(\alpha^v). \tag{16}$$

The sufficient condition for proof of an ascent direction (12) for the Armijo iteration is hence trivially satisfied without modification.

IV. SOME COMPUTATIONAL EXPERIENCE

Three examples were created for testing the algorithm on the University of Virginia's CDC Cyber 855. Samples of size 250 were generated as random samples according to each of three assumed densities, namely,

(1) $g(t) = .5(e^{-t}) + .25(3e^{-3t}) + .25(5e^{-5t})$;

(2) $g(t) = 4(2e^{-2t}) - 4(3e^{-3t}) + 1(4e^{-4t})$;

(3) $g(t) = 2(2e^{-2t}) - 2(4e^{-4t}) + 1(6e^{-6t})$.

Plots of these densities follow in Figures 1-3; the particular forms were chosen strictly for illustration. The first one is a simple example of a completely monotone, ordinary exponential mixture, while the second one is monotone decreasing but with two points of inflection. The third case, on the other hand, is a unimodal density with shape determined by the relative size of its negative middle γ-value.

The random variates for the first density were easily created in the usual composition or mixture way. Given the descending order of the scale parameters (say $\eta_i = 1/\phi_i$) 1, 1/3, 1/5, we know that the optimization routine should find a unique solution, for the algorithm collapses to the K/M/H method in that case. Two test subcases were run, starting from equal sets of mixture probabilities but with very different scale parameters. The algorithm came up with nearly the same answer in almost equal time. Note that although the algorithm required 98 and 97 iterations, these are done very quickly, with the total run requiring less than 10.4 seconds cpu time.

The salient results of the two runs for test case number 1 are displayed in Table 1, and the estimated density resulting from the first starting point is plotted in Figure 4. Though the simulated data set led to different

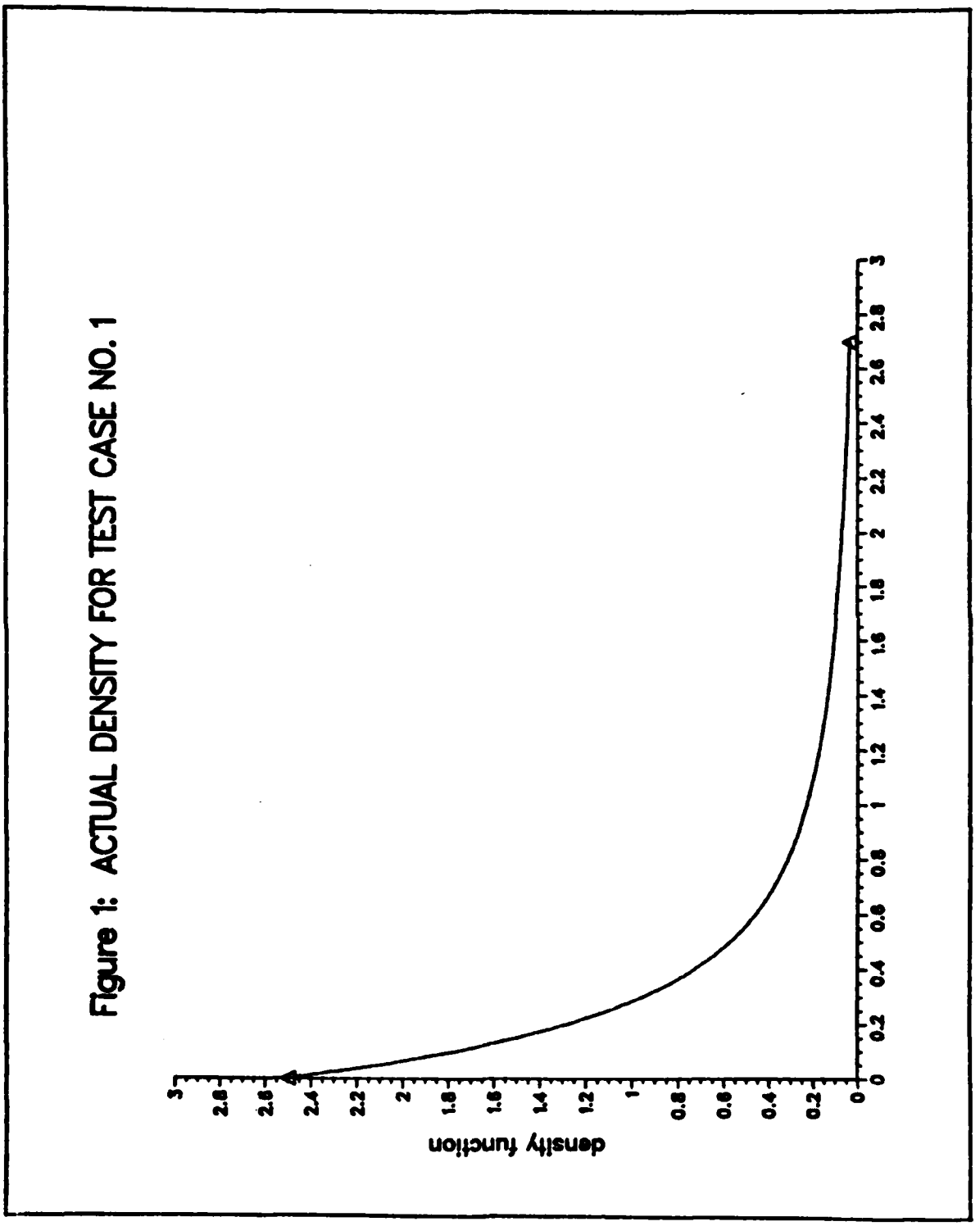


Figure 1: ACTUAL DENSITY FOR TEST CASE NO. 1

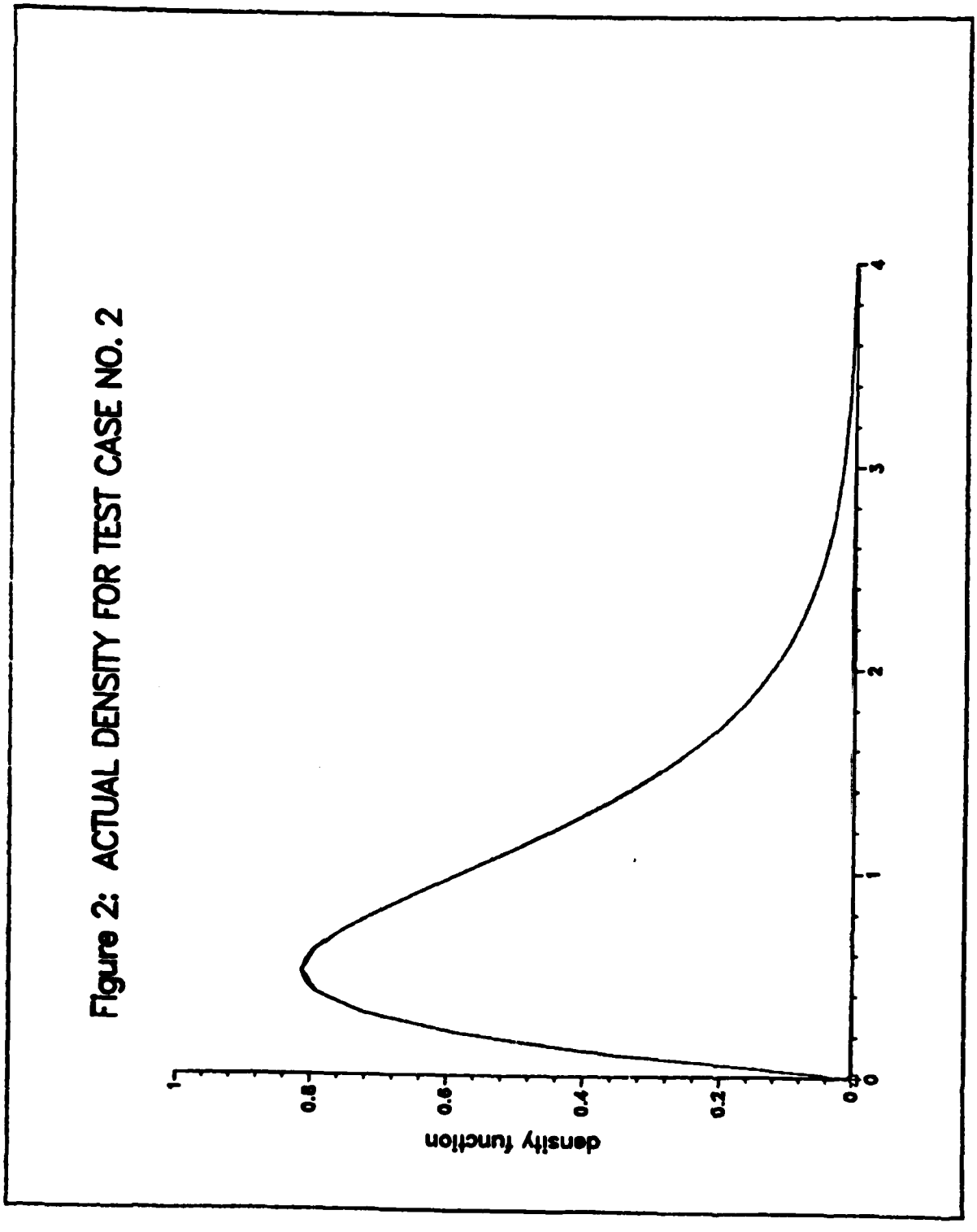


Figure 2: ACTUAL DENSITY FOR TEST CASE NO. 2

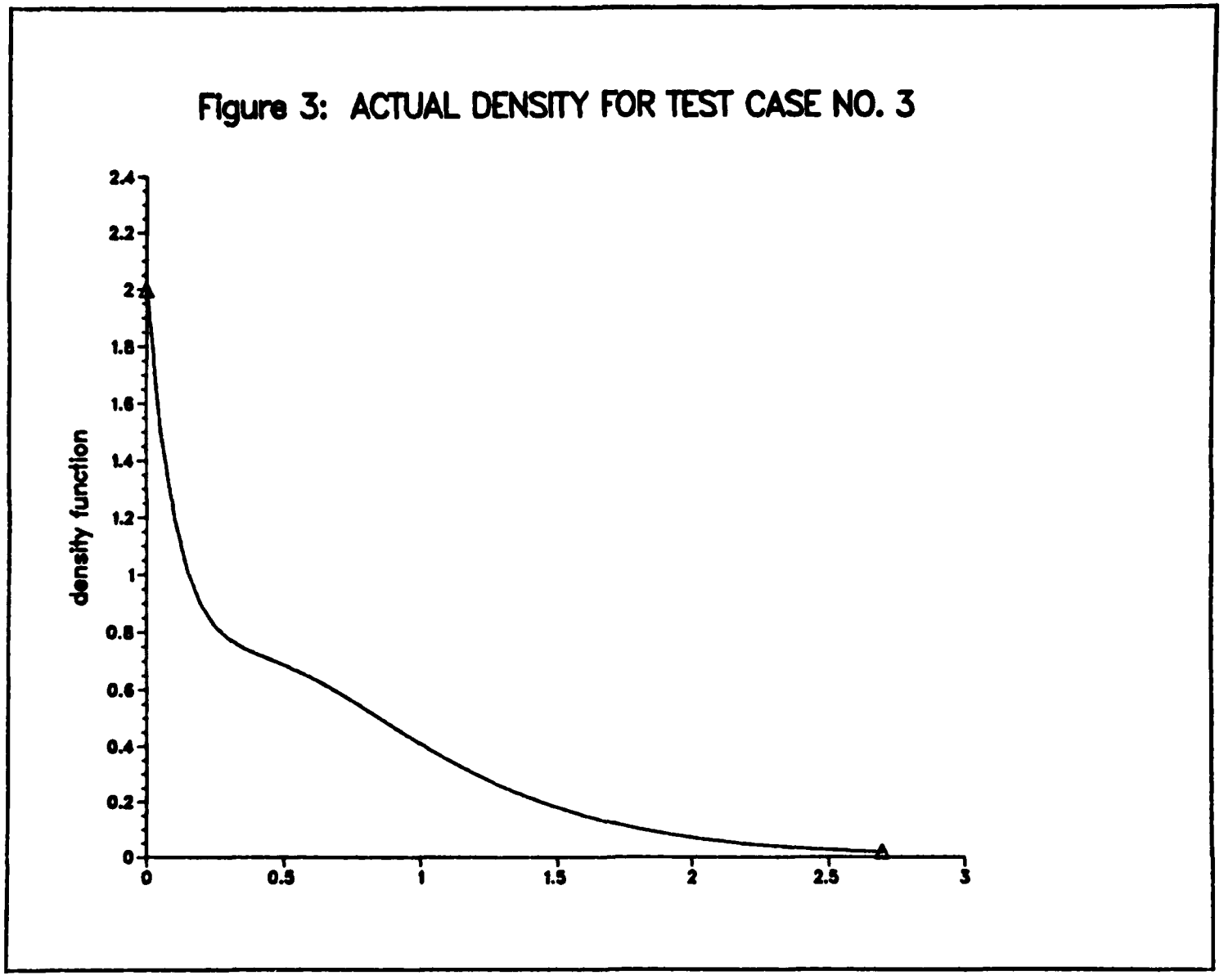


Figure 3: ACTUAL DENSITY FOR TEST CASE NO. 3
density function
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
2
2.5
3

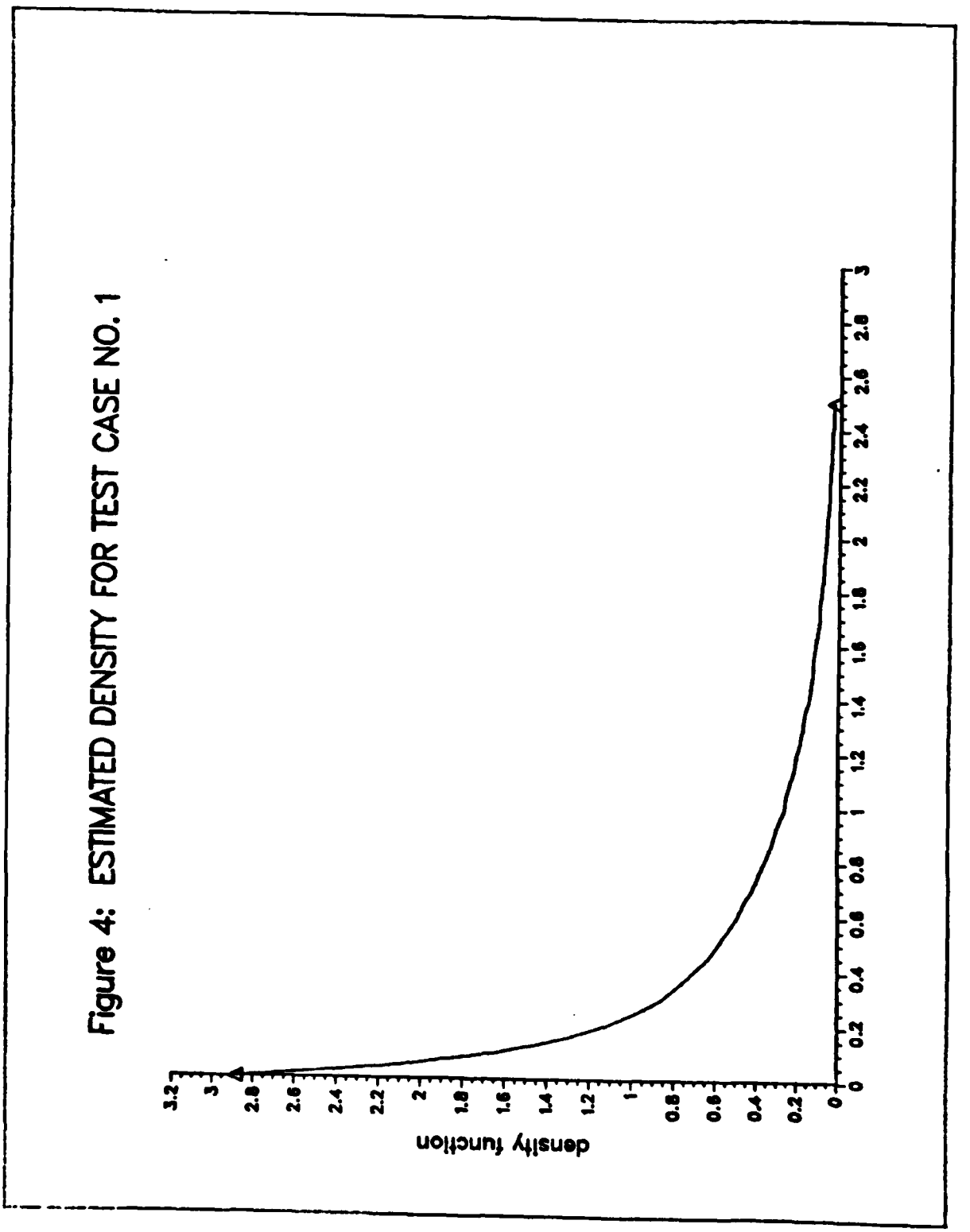


Figure 4: ESTIMATED DENSITY FOR TEST CASE NO. 1

estimates than the original values, we know that the respective values would eventually get closer with increased sample sizes. But the most important point here is that the algorithm worked well.

Table 1

TEST CASE NUMBER 1

Actual density $g(t) = \frac{1}{2} e^{-t} + \frac{3}{4} e^{-3t} + \frac{5}{4} e^{-5t}$

with E[T] = 19/30 = $.6\bar{3}$
and Var[T] = 607/900 = $.67\bar{4}$.

	(=1) η_1	(=$.\bar{3}$) η_2	(=.2) η_3	(=.5) γ_1	(=.25) γ_2	(=.25) γ_3	objective function
(i) starting point	.5	.4	.3	.3	.3	.3	-151.10
answer	.90	.51	.09	.44	.41	.15	-122.36
	(mean = .6174, variance = .4598)						
cpu seconds	10.365						
# iterations	98						
# bisections required	14: 9 for negative η, 5 for decrement in objective function						
(ii) starting point	1.0	.5	.25	.3	.3	.3	-124.43
answer	.90	.51	.09	.42	.43	.15	-122.35
	(mean = .6174, variance = .4605)						
cpu seconds	10.368						
# iterations	97						
# bisections required	14: 9 for negative η, 5 for decrement in objective function						

The random variates for the second and third cases could not be created by composition because of the negative mixing parameters. Instead, each of these two densities was rewritten in mixed generalized Erlang form, from which the variates were easily derived using composition. When written in convolution format, these specific equivalences turned out to be

$$g(t) = \frac{2}{3}\,(2e^{-2t} * 3e^{-3t} * 4e^{-4t}) + \frac{1}{3}\,(3e^{-3t} * 4e^{-4t}) \qquad \text{(Case 2)}$$

and

$$g(t) = \frac{2}{3}\,(2e^{-2t} * 4e^{-4t} * 6e^{-6t}) + \frac{1}{3}\,(6e^{-6t}) \qquad \text{(Case 3)}$$

(The equalities are verified in the simplest way by converting to Laplace transforms.)

For these two test cases, we ran from three different starting points, one of which was the actual answer, while another was chosen to be "close". The results are displayed in Tables 2 and 3.

For the second case, we observe that the first and third starting points led to answers which look like they may well be the same (and close to the actual). There are differences in values, but we theorize that the first answer would move toward the third if the algorithm's stopping rule were tightened. But it would appear that the second starting point did lead to a truly alternative local solution. For purposes of comparison, in Figure 5, we offer a plot of the density functions which result from the first two solutions. We see there that the two unimodal densities are relatively close to each other.

For this second test case, note that cpu seconds ranged from a high of 13.8 seconds to a low of 4.8. Through all three subcases, the relative use of

Table 2

TEST CASE NUMBER 2

Actual density $g(t) = 4(2e^{-2t}) - 4(3e^{-3t}) + 1(4e^{-4t})$

with $E[T] = .91\overline{6}$
and $Var[T] \doteq .3544$

		(=.5) η_1	(=.$\overline{3}$) η_2	(=.25) η_3	(=4) γ_1	(=-4) γ_2	(=1) γ_3	objective function
(i)	starting point	.5	.4	.3	5	-5	1	-208.28
	answer	.50	.38	.32	4.96	-4.94	0.98	-197.14
		(mean = .9072, variance = .4268)						
	cpu seconds	13.766						
	# iterations	41						
	# bisections required	163: 5 for negative η, 127 for decrement in objective, 31 for nondensity appearance						
(ii)	starting point	.5	.4	.3	1	-1	1	-332.16
	answer	.71	.55	.27	.97	.75	-.72	-198.54
		(mean = .9001, variance = .5013)						
	cpu seconds	7.891						
	# iterations	75						
	# bisections required	75: 1 for negative η, 65 for decrement in objective, 9 for nondensity appearance						
(iii)	starting point	.5	.$\overline{3}$	.25	4	-4	1	-198.14
						(actual answer)		
	answer	.50	.34	.25	4.04	-4.05	1.02	-197.00
		(mean = .9000, variance = .3999)						
	cpu seconds	4.785						
	# iterations	13						
	# bisections required	49: 45 for decrement in objective, 4 for nondensity appearance						

Table 3

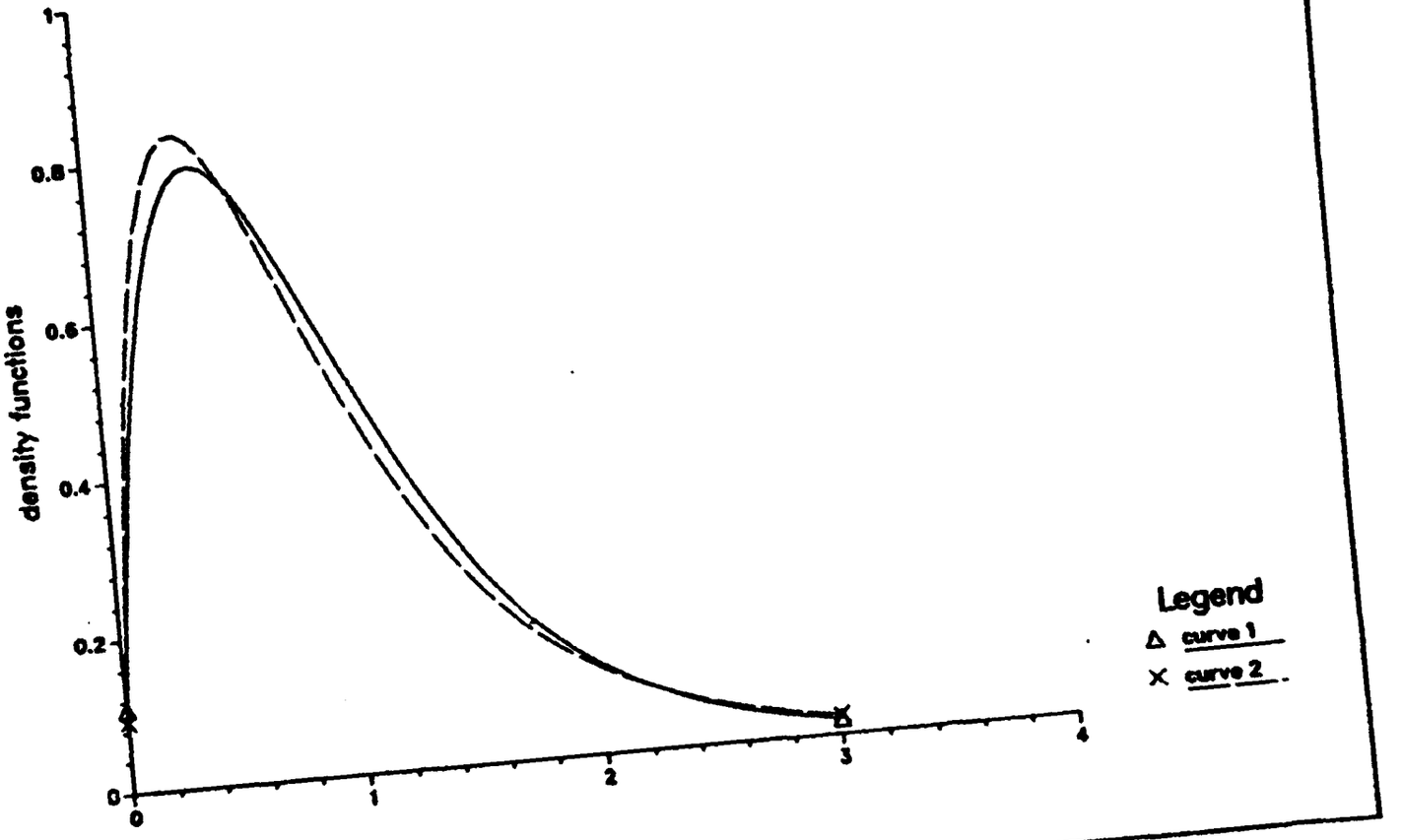
TEST CASE NUMBER 3

Actual density $g(t) = 2(2e^{-2t}) - 2(4e^{-4t}) + 1(6e^{-6t})$

with $E[T] = .\bar{6}$
and $Var[T] = .36\bar{1}$

		(=.5) η_1	(=.25) η_2	(=.1$\bar{6}$) η_3	(=2) γ_1	(=-2) γ_2	(=1) γ_3	objective function
(i)	starting point	.5	.4	.3	5	-5	1	-146.06
	answer	.42	.34	.20	5.09	-5.10	1.01	-134.74
		(mean = .6408, variance = .3728)						
	cpu seconds	71.743						
	# iterations	307						
	# bisections required	692: 44 for negative η, 617 for decrement in objective, 31 for nondensity appearance						
(ii)	starting point	.5	.4	.3	1	-1	1	-170.19
	answer	.60	.30	.29	1.10	1.00	-1.10	-137.14
		(mean = .6363, variance = .3905)						
	cpu seconds	2.605						
	# iterations	7						
	# bisections required	24: 23 for decrement in objective, 1 for nondensity appearance						
(iii)	starting point	.5	.25	.1$\bar{6}$	2	-2	1	-135.11
						(actual answer)		
	answer	.48	.25	.17	2.04	-2.06	1.03	-134.70
		(mean = .6381, variance = .3705)						
	cpu seconds	3.206						
	# iterations	18						
	# bisections required	15: 15 for decrement in objective						

Figure 5: ALTERNATIVE DENSITIES FOR TEST CASE NO. 2



the bisection rule was comparable, most frequently resulting from a non-improvement in the objective function.

Of course, it should be understood that in this work we offer no definitive strategies for locating the global solutions from amongst the locals. The ability to arrive at the global solution likely depends on the quality of the starting point. In these test cases, all starting points were selected in an arbitrary fashion. Instead, we could use a moment matching technique for the initial point, or possibly even a numerical curve fitting technique (with parameters adjusted up or down to make sure it is a density). Two such numerical approaches are documented in McDonough and Huggins (1968) and Harman and Fairman (1973).

The results for our third test cases appear somewhat different in the sense that we seem to have found three local solutions. For comparison here, we plotted the first two experimental densities in Figure 6 and again see that they are close to each other. The third one is also quite similar, but we opted not to plot it, noting instead that it is almost identical to the originating density as plotted in Figure 3. This time, cpu seconds ranged from 2.6 to 71.7. Again, there is a preponderance of bisection because of non-improvement, though negative η parameters caused 44 out of 692 bisections in subcase i. Overall, the algorithm again performed well.

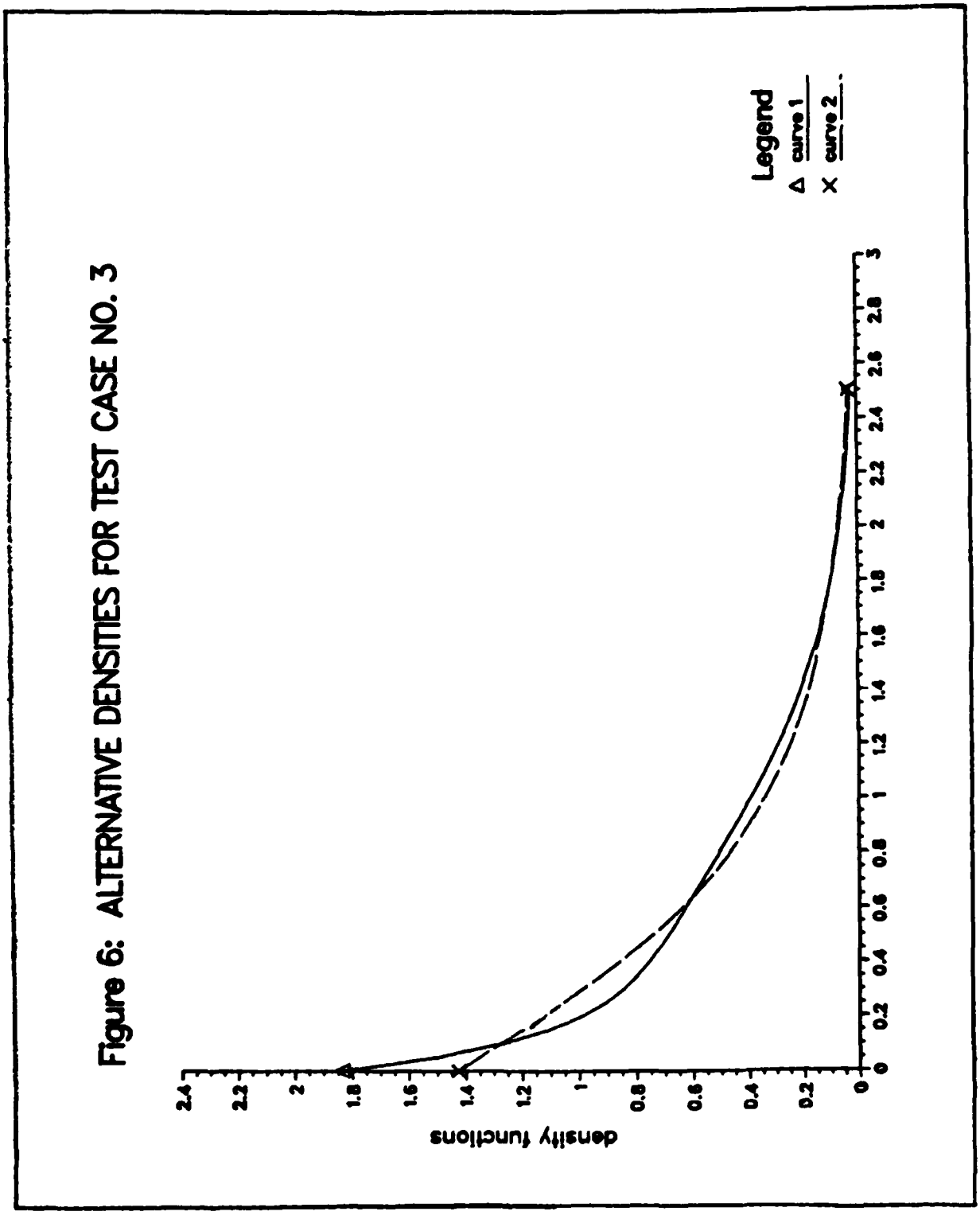


Figure 6: ALTERNATIVE DENSITIES FOR TEST CASE NO. 3

V. CONCLUDING REMARKS

To close, we repeat our observation that the algorithm worked well. It is clearly also well suited for the MLE of parameters from ordinary exponential mixtures since it is guaranteed then to find the global solution.

Primary areas of possible future work include an exploration of the statistical properties of these estimators, a firmer strategy for selecting "good starting points", and the derivation of possible procedures for determining the optimum number of terms to include. This last concern is akin to the step-wise regression problem and is often mentioned in the literature as a topic of special interest. Our current code can handle any number of terms, but that number must be specified beforehand.

REFERENCES

L. Armijo, "Minimization of Functions Having Lipschitz Continuous First Partial Derivatives," Pacific Journal of Mathematics, Vol. 16, No. 1 (1966), pp. 1-3.

D. R. Cox, "A Use of Complex Probabilities in the Theory of Stochastic Processes," Proc. Camb. Phil. Soc., Vol. 51 (1955), pp. 313-319.

M. Dehon and G. Latouche, "A Geometric Interpretation of the Relations Between the Exponential and Generalized Erlang Distributions," Adv. Appl. Prob., Vol. 14 (1982), pp. 885-897.

B. S. Everitt and D. J. Hand, Finite Mixture Distributions, London: Chapman and Hall, 1981.

D. Gross and C. M. Harris, Fundamentals of Queueing Theory, New York: John Wiley, 1974.

R. K. Harman and F. W. Fairman, Exponential Approximation via a Closed Form Gauss-Newton Method, IEEE Trans. Circuit Theory, CT-20 (1973), pp. 361-369.

A. R. Kaylan, "Statistical Analysis of Finite Mixtures of Exponential and Weibull Mixtures," Ph.D. Dissertation, Department of Industrial Engineering and Operations Research, Syracuse University, Syracuse, New York, 1978.

A. R. Kaylan and C. M. Harris, "Efficient Algorithms to Derive Maximum-Likelihood Estimates for Finite Exponential and Weibull Mixtures," Computers and Operations Research, Vol. 8 (1981), pp. 97-104.

J. Mandelbaum, "Numerical Issues in Inference for Finite Mixtures of Weibull Distributions Under Progressive Censoring," Ph.D. Dissertation, Department of Operations Research, The George Washington University, Washington, D.C., 1982.

J. Mandelbaum and C. M. Harris, "Parameter Estimation Under Progressive Censoring Conditions for a Finite Mixture of Weibull Distributions," Studies in the Management Sciences, Vol. 19 (1982), pp. 239-260.

R. N. McDonough and W. H. Huggins, Best Least Squares Representation of Signals by Exponentials, IEEE Trans. Autom. Control, AC-13 (1968), pp. 408-412.

A. W. Naylor and G. R. Sell, Linear Operator Theory in Engineering and Science, New York: Holt, Rinehart and Winston, 1971.

M. F. Neuts, Matrix-Geometric Solutions in Stochastic Models, Baltimore: Johns Hopkins, 1981.

J. M. Ortega and W. C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, New York: Academic Press, 1970.

W. L. Smith, "On the Distribution of Queueing Times," Proc. Camb. Phil. Soc., Vol. 49 (1953), pp. 449-461.

4650:cr

DISTRIBUTION LIST

Copy No.

1 Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Attention: Group Leader, Statistics and Probability
Mathematical and Physical Sciences

2 Office of Naval Research Resident Representative
Joseph Henry Bldg., Rm. 623
2100 Pennsylvania Avenue., NW
Washington, DC 20037

3 - 9 Director
Naval Research Laboratory
Washington, DC 20375

Attention: Code 2627

10 - 22 Defense Technical Information Center
Bldg. 5, Cameron Station
Alexandria, VA 22314

23 Mr. Seymour Selig
Statistics and Probability Program
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

24 - 25 C. M. Harris

26 - 27 E. H. Pancake, Clark Hall

28 RLES Files

JO#4950:cr

UNIVERSITY OF VIRGINIA
School of Engineering and Applied Science

The University of Virginia's School of Engineering and Applied Science has an undergraduate enrollment of approximately 1,500 students with a graduate enrollment of approximately 500. There are 125 faculty members, a majority of whom conduct research in addition to teaching.

Research is a vital part of the educational program and interests parallel academic specialties. These range from the classical engineering disciplines of Chemical, Civil, Electrical, and Mechanical and Aerospace to newer, more specialized fields of Biomedical Engineering, Systems Engineering, Materials Science, Nuclear Engineering and Engineering Physics, Applied Mathematics and Computer Science. Within these disciplines there are well equipped laboratories for conducting highly specialized research. All departments offer the doctorate; Biomedical and Materials Science grant only graduate degrees. In addition, courses in the humanities are offered within the School.

The University of Virginia (which includes approximately 1,500 full-time faculty and a total full-time student enrollment of about 16,000), also offers professional degrees under the schools of Architecture, Law, Medicine, Nursing, Commerce, Business Administration, and Education. In addition, the College of Arts and Sciences houses departments of Mathematics, Physics, Chemistry and others relevant to the engineering research program. The School of Engineering and Applied Science is an integral part of this University community which provides opportunities for interdisciplinary work in pursuit of the basic goals of education, research, and public service.

END

FILMED

10-84

DTIC